



Aprendizado de Máquina
Introdução às Redes Neurais
Artificiais

Marcos Oliveira Prates
(Agradecimento Marcelo Azevedo Costa)
Departamento de Estatística
Universidade Federal de Minas Gerais

Inteligência Artificial (IA)

O objetivo da inteligência artificial é o desenvolvimento de paradigmas ou algoritmos que requeiram máquinas para realizar tarefas cognitivas, para as quais os humanos são atualmente melhores. (Sage, 1990)

Um sistema de IA deve ser capaz de fazer três coisas:

1. Armazenar conhecimento
 2. Aplicar o conhecimento armazenado para resolver problemas
 3. Adquirir novo conhecimento através da experiência
-

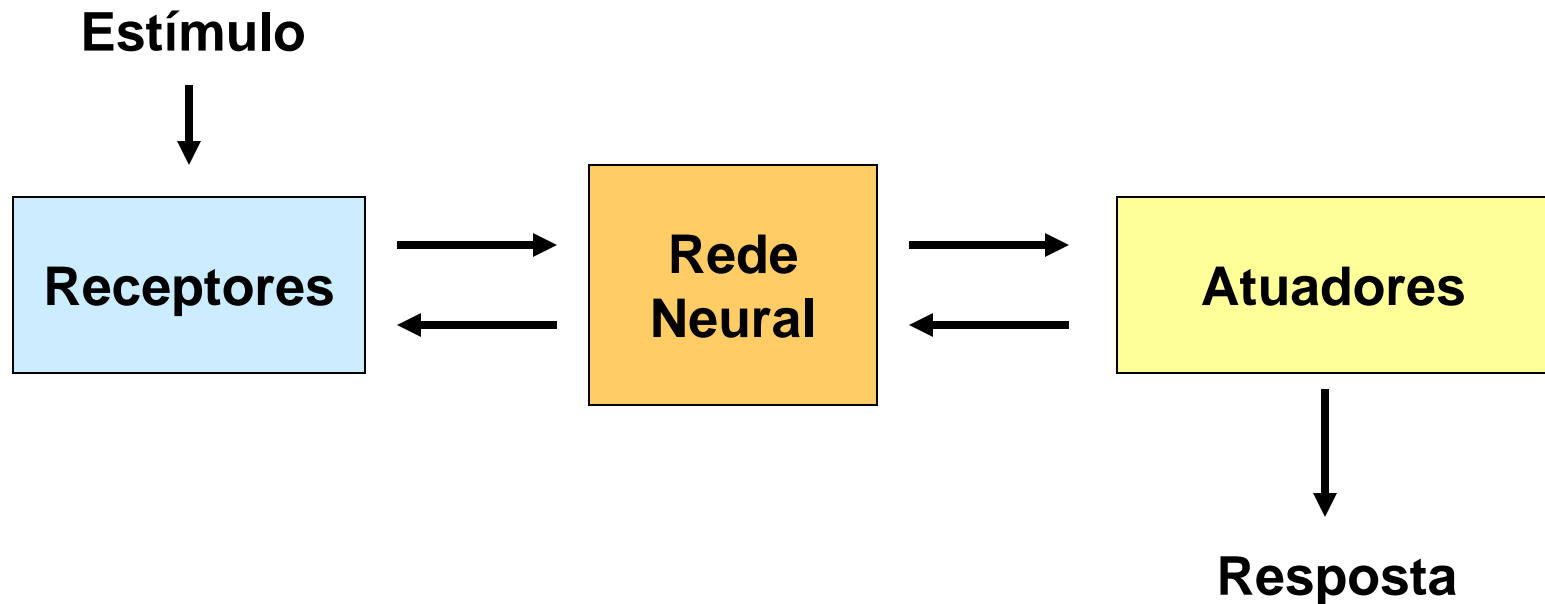
Motivação para as RNAs

- O cérebro contém 10^{11} neurônios interconectados e operando em paralelo.
 - O cérebro é responsável pela emoção, pensamento, percepção, cognição, execução de funções sensoriomotorase autômatas.
 - O funcionamento das redes biológicas não é amplamente conhecida.
 - Tem-se modelos simplificados de estruturas fisiológicas básicas.
-

Histórico das Redes Neurais Artificiais

- A primeira Rede Neural foi apresentada por Alexander Bain em 1873 no livro intitulado “*Mind and Body. The Theories of Their Relation*”
 - O trabalho consistia em pesquisa na área de neuroanatomia
 - Warren McCulloch (psiquiatra e neuroanatomista) e Walter Pitts (matemático) em 1943 apresentaram o primeiro modelo artificial de um neurônio biológico (nodos MCP): “A Logical Calculus of the Ideas Immament in Nervous Activity”
-

Representação em Diagrama de Blocos do Sistema Nervoso



Por que é interessante propor um modelo para o comportamento humano?

Aprendizado

Tarefa: Aprender a andar



Inteligência Artificial



Sumário

- Visão geral (Redes Biológicas)
 - O Neurônio biológico
 - O Modelo Matemático
 - Redes Multi-Camadas
 - **Redes Auto-Organizativas**
 - **Aplicações**
-

O que são Redes Neurais Artificiais

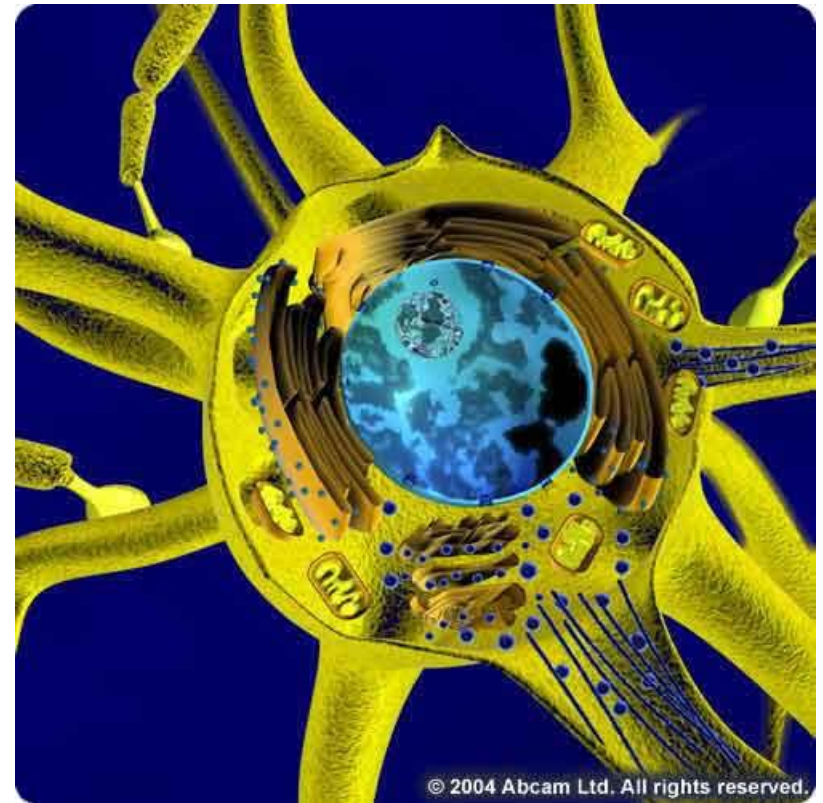
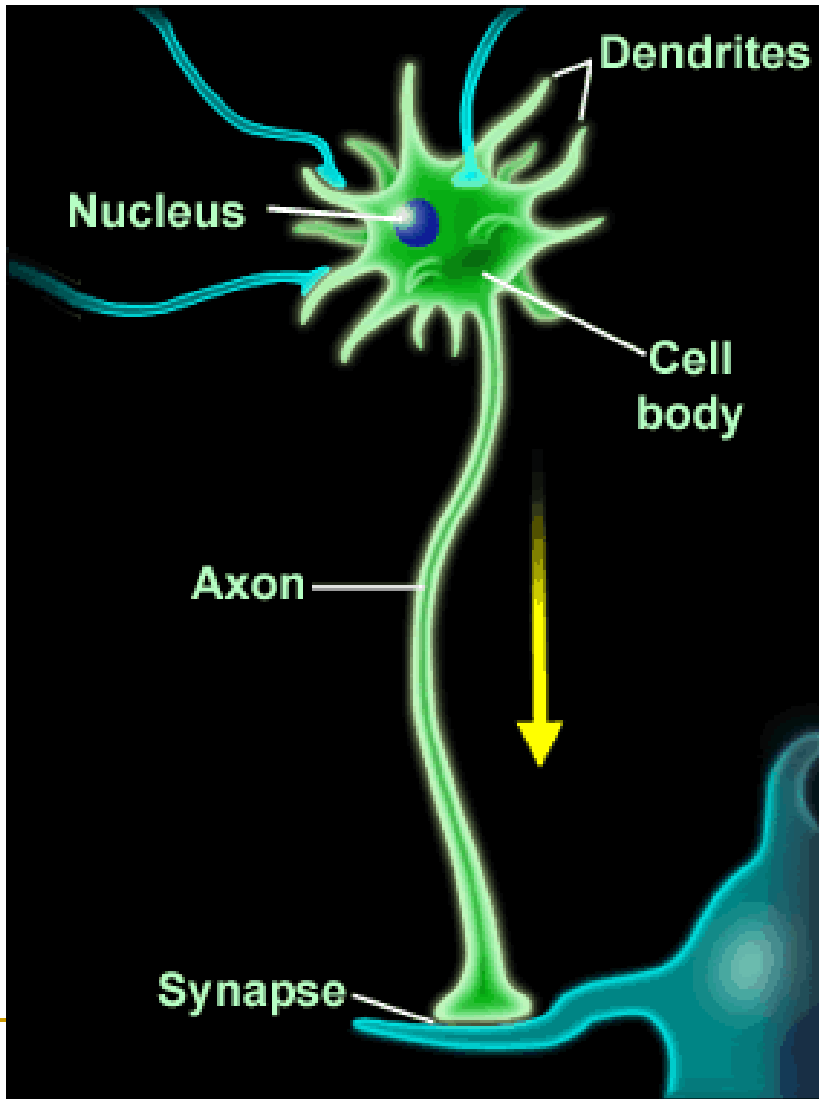
- Redes Neurais Artificiais (RNAs) são tentativas de produzir sistemas de aprendizado biologicamente realísticos
 - São baseados em modelos abstratos de como *pensamos* que o cérebro (e os neurônios) funcionam
 - Compostas por várias unidades de processamento (*neurônios*)
 - Interligadas por um grande número de conexões
-

O que são Redes Neurais Artificiais

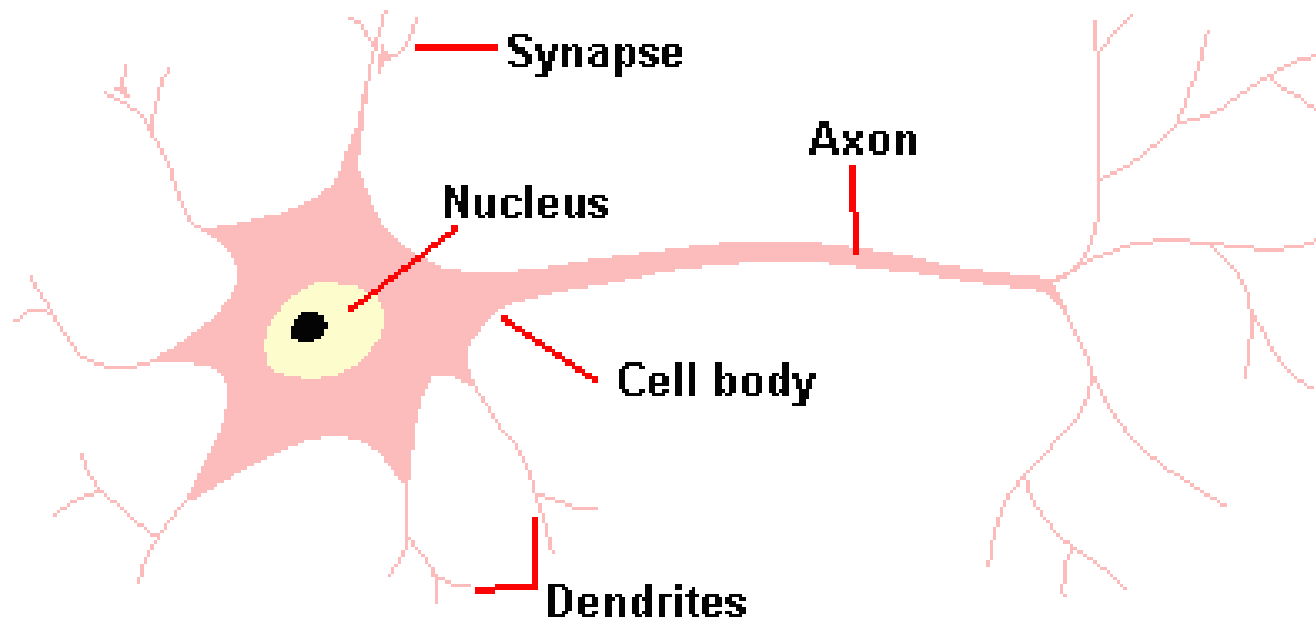
- Redes Neurais Artificiais (RNAs) são modelos de computação com propriedades particulares:
- Capacidade de se adaptar ou aprender através de exemplos
- Capacidade de generalizar
- Agrupar ou organizar dados
- Tolerância a falhas
- Implementação rápida

Multi-disciplinaridade: Ciência da Computação, Matemática, Física, Engenharias, Psicologia, Biologia, Linguística, Filosofia, etc.

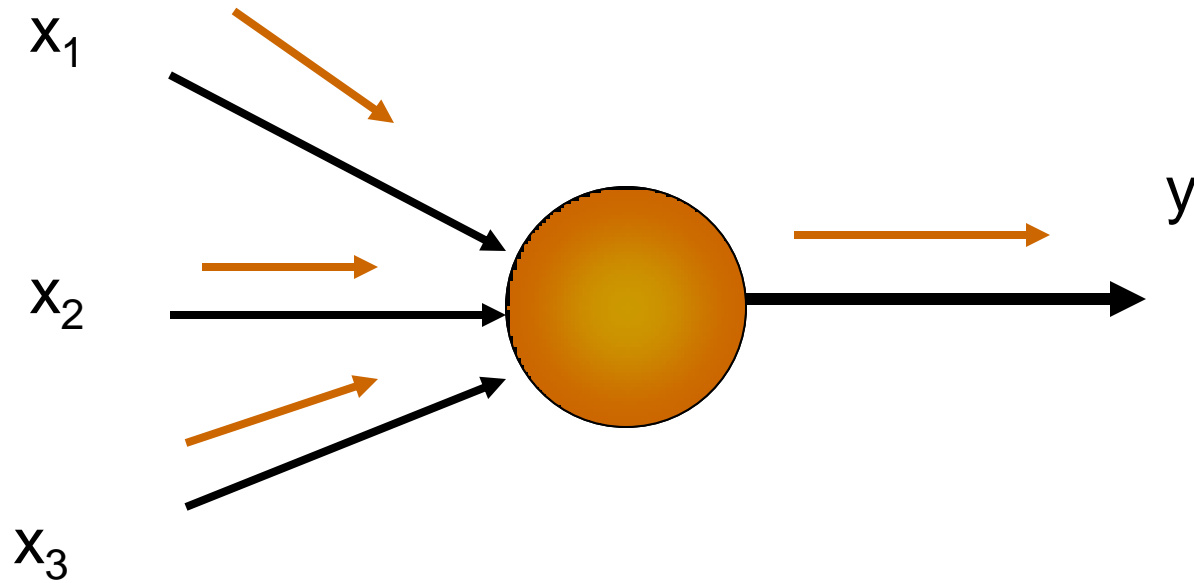
O Neurônio Biológico



O Neurônio Biológico



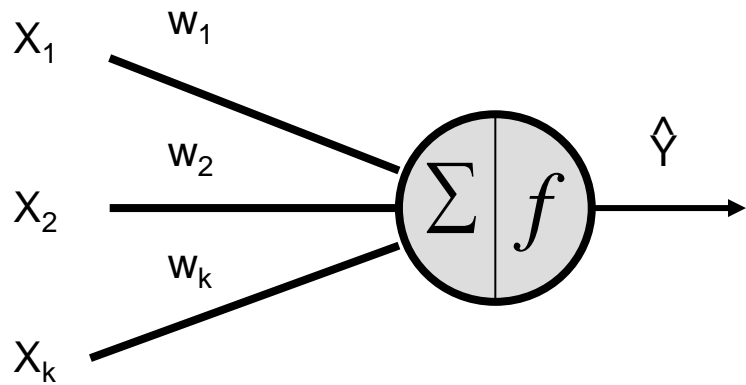
O Neurônio Biológico



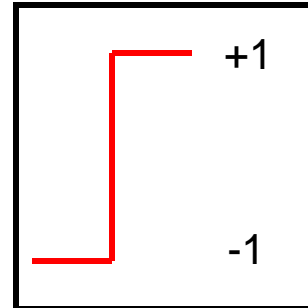
$$y = f(x_1, x_2, x_3)$$

$$(x_1, x_2, x_3, y) \in \mathfrak{R}$$

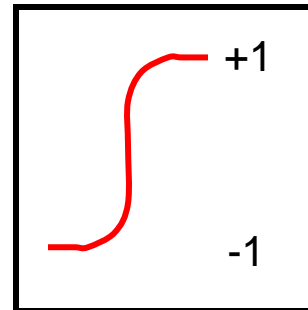
O Nodo Perceptron



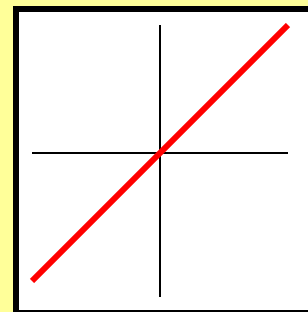
$$\hat{Y} = f\left(\sum_{i=1}^k x_i \cdot w_i\right)$$



$$\hat{Y} = \text{sgn}\left(\sum_{i=1}^k x_i \cdot w_i\right)$$

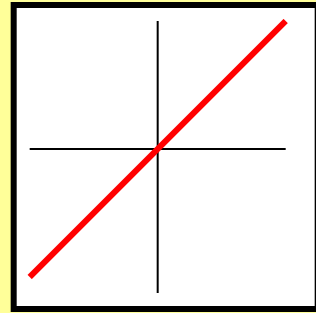


$$\hat{Y} = \tanh\left(\sum_{i=1}^k x_i \cdot w_i\right)$$



$$\hat{Y} = \sum_{i=1}^k x_i \cdot w_i$$

Modelos Estatísticos Lineares



$$\hat{Y} = \sum_{i=1}^k x_i \cdot w_i$$

$$\hat{y} = \sum_{i=1}^k x_i \cdot \hat{\beta}_i$$

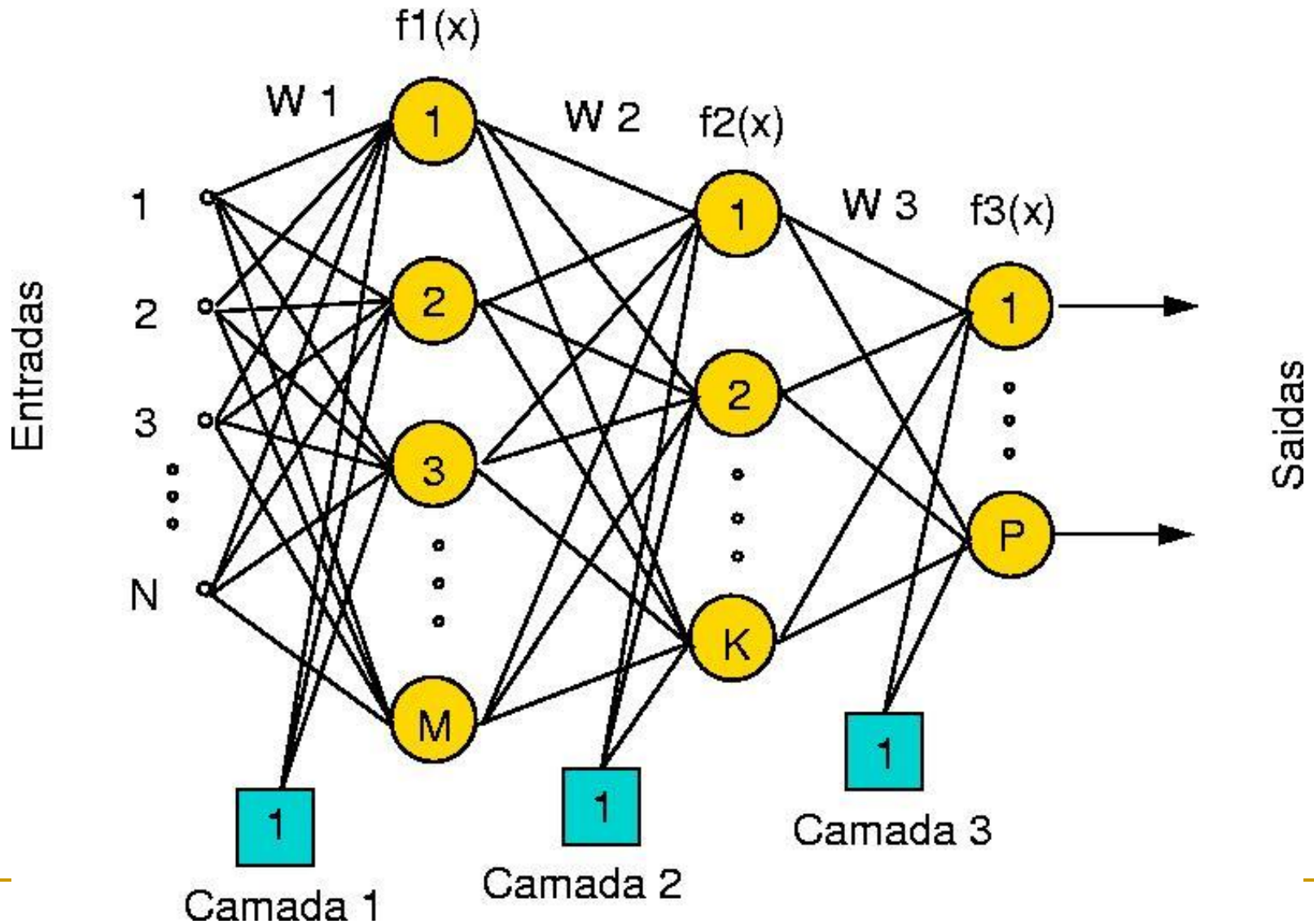
Modelo de regressão linear

$$\hat{y}_t = \hat{\phi}_1 \cdot y_{t-1} + \hat{\phi}_2 \cdot y_{t-2}$$

Modelo Auto Regressivo de ordem 2

$$\{x \mid \beta_0 + \beta^T x = 0\} \text{ Hiperplano de separação entre duas classes}$$

Rede *Multi-Layer Perceptron* (MLP)



Exemplo de Regressão Linear

j	x_j	y_j
1	-1.0000	-1.0000
2	-0.5000	0.5000
3	0	2.0000
4	0.5000	3.5000
5	1.0000	5.0000
6	1.5000	6.5000
7	2.0000	8.0000
8	3.0000	11.0000

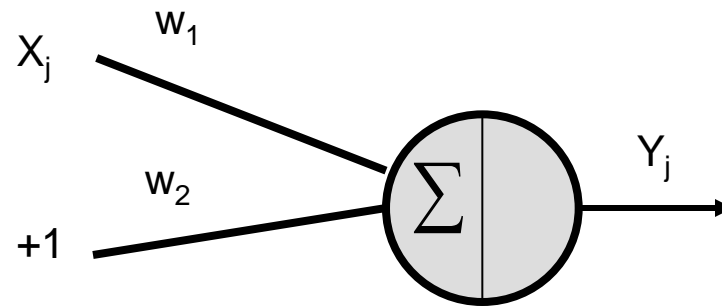
$$y_j = 3.x_j + 2$$

Modelo de Regressão Linear

$$y_j = \hat{\beta}_1 . x_j + \hat{\beta}_2$$

Modelo Neural

$$y_j = \omega_1 . x_j + \omega_2$$



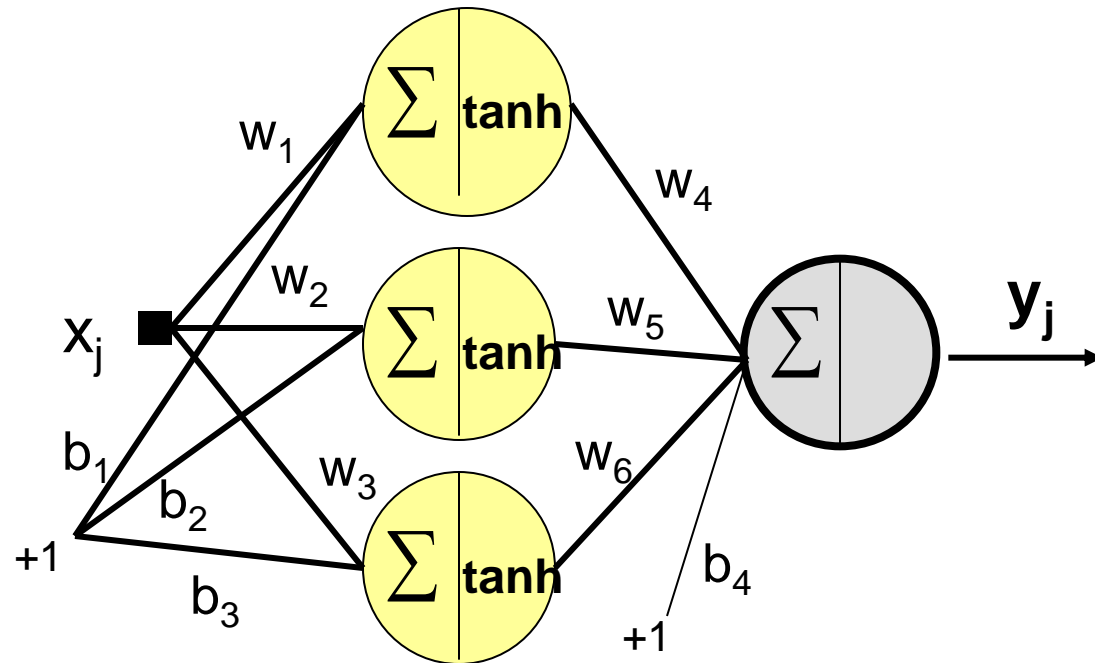
Exemplo de Regressão não - Linear

Modelo Polinômial

$$y_j = \hat{\beta}_1 \cdot x_j^2 + \hat{\beta}_2 \cdot x_j + \hat{\beta}_3$$

j	x_j	y_j
1	-1.0000	2.0000
2	-0.5000	0.7500
3	0	1.0000
4	0.5000	2.7500
5	1.0000	6.0000
6	1.5000	10.7500
7	2.0000	17.0000
8	2.5000	24.7500

Modelo Neural



$$y_j = 3 \cdot x_j^2 + 2 \cdot x_j + 1$$

$$y_j = \omega_4 \cdot \tanh(\omega_1 \cdot x_j + b_1) + \omega_5 \cdot \tanh(\omega_2 \cdot x_j + b_2) + \omega_6 \cdot \tanh(\omega_3 \cdot x_j + b_3) + b_4$$

Exemplo de Regressão

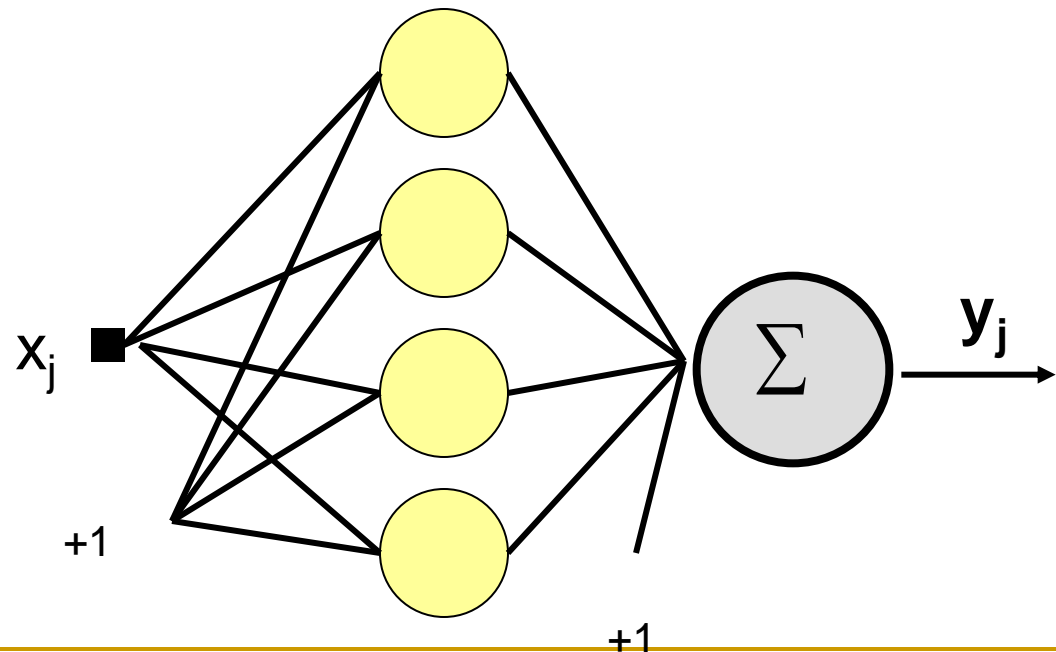
Modelo Polinômial

$$y_j = \hat{\beta}_n \cdot x_j^n + \hat{\beta}_{n-1} \cdot x_j^{n-1} + \dots + \hat{\beta}_0$$

j	x_j	y_j
1	-1.0000	-0.6031
2	-0.5000	-0.6915
3	0	0
4	0.5000	1.1915
5	1.0000	2.6031
6	1.5000	4.1526
7	2.0000	5.8733
8	2.5000	7.8351
9	3.0000	10.1362

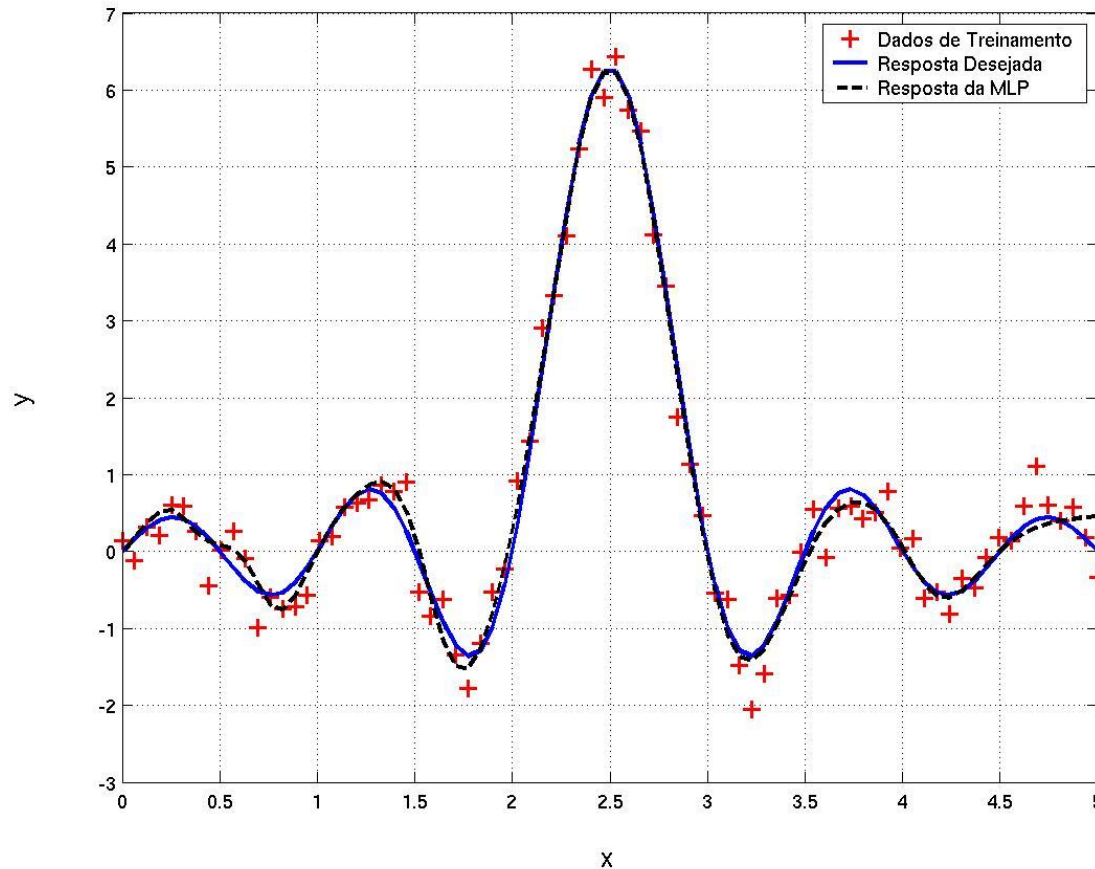
?

Modelo Neural



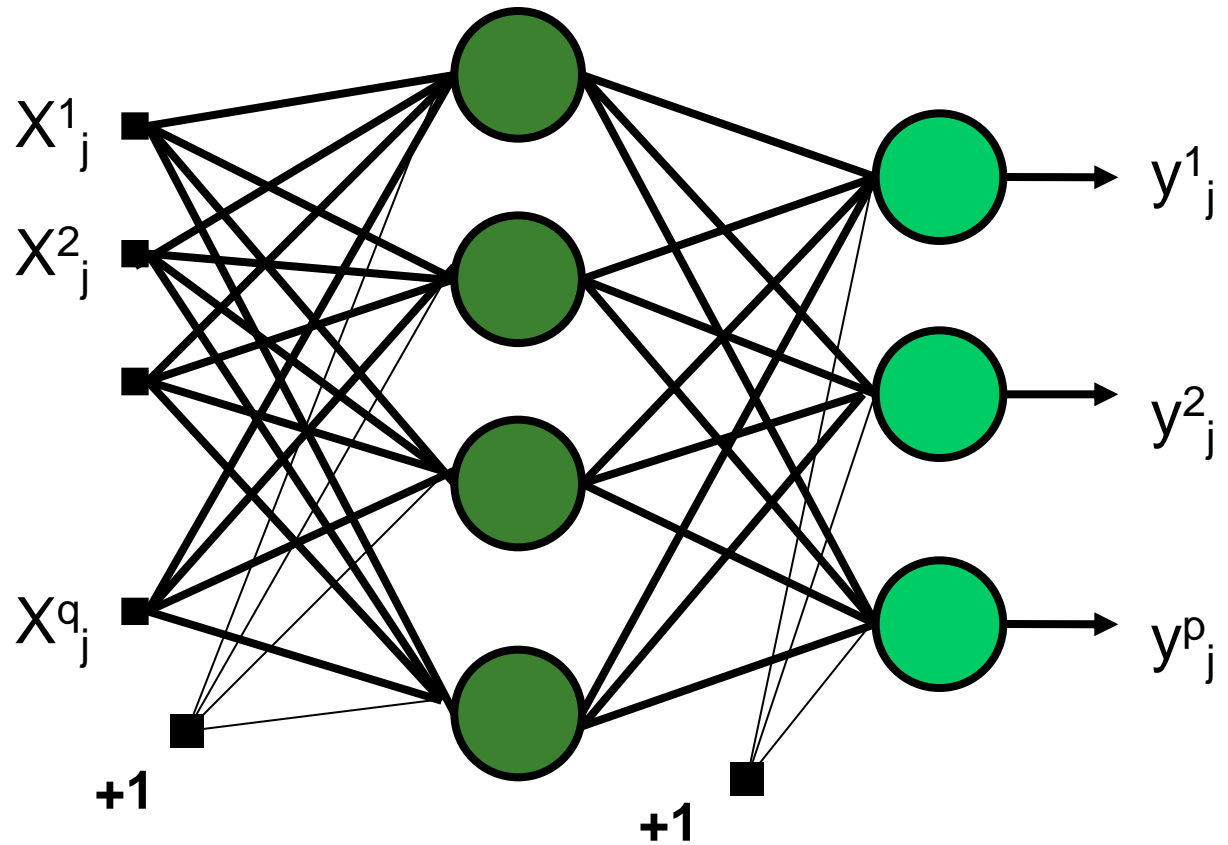
É um modelo flexível !!!

Exemplo de aproximação de função utilizando uma rede MLP com 20 nodos na camada escondida - O equilíbrio entre o erro e a norma garante modelos com alta capacidade de generalização independente do número de parâmetros (variáveis de ajuste) do modelo.



Exemplo de Regressão Multivariada

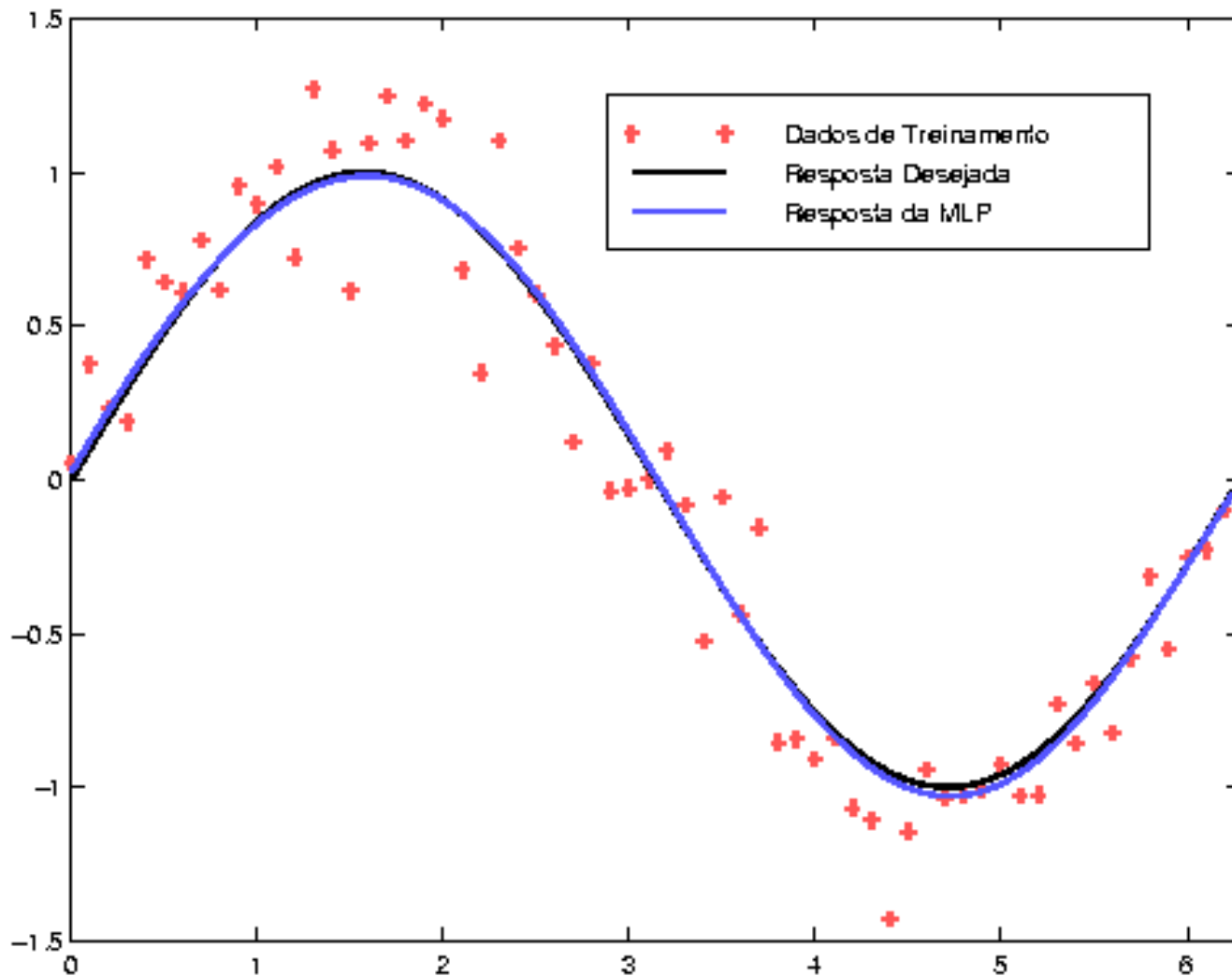
x_j	y_j
x_1	y_1
x_2	y_2
x_3	y_3
x_n	y_n



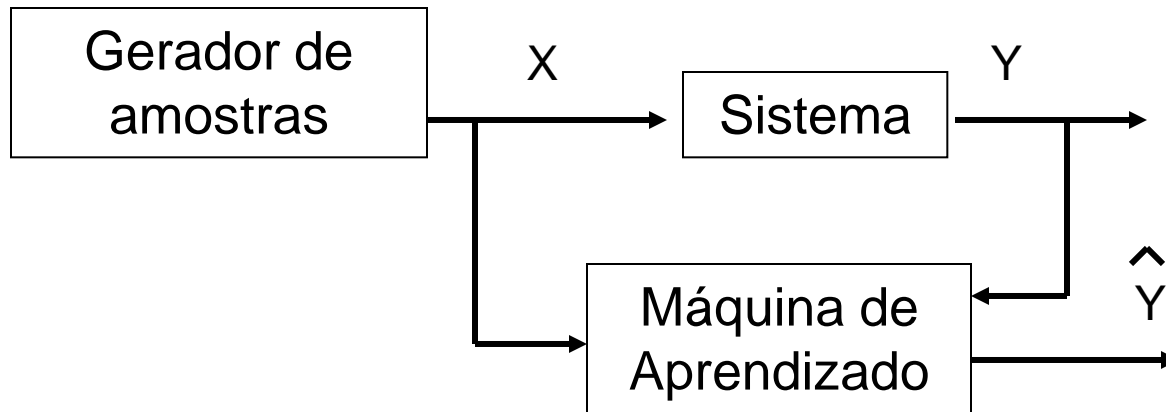
$$\mathbf{X}_j = [x_j^1, x_j^2, x_j^3, \dots, x_j^q]^T$$

$$\mathbf{y}_j = [y_j^1, y_j^2, y_j^3, \dots, y_j^p]^T$$

Exemplo da Flexibilidade de uma RNA (MLP)



O Modelo Geral de Aprendizagem Supervisionada



$$y = g(x) + \varepsilon \quad \hat{y} = f(x, w) \quad w \in \mathfrak{R}^N$$

Conjunto de observações: $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

Objetivo: $f(x, w) \approx g(x)$ **(Generalização)**

- A máquina de aprendizado deve ser capaz de prever respostas coerentes em relação às realizações futuras de x .
- A partir da máquina de aprendizado, deve ser possível extrair informações sobre a natureza das relações entre as variáveis de entrada e saída.

Treinamento Supervisionado de uma Rede Neural Artificial

Função de Custo

$$J(w) = \frac{1}{2} \sum_{j=1}^n (y_j - f(x_j, w))^2$$

Objetivo: Minimizar a função de custo em relação a um conjunto de observações

Seja agora, um vetor de parâmetros:

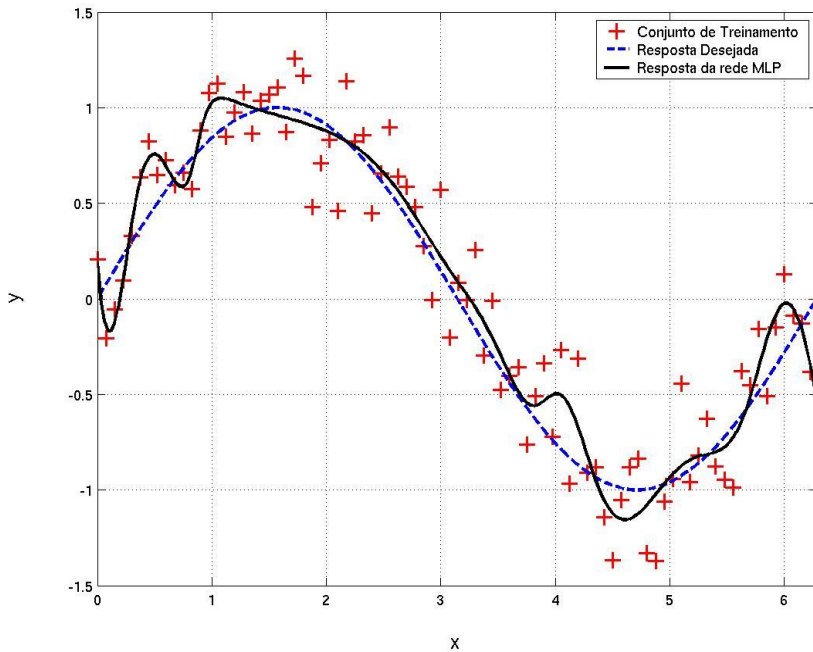
$$\mathbf{w} = [w_1, w_2, w_3, \dots, w_r, b_1, b_2, \dots, b_t]$$

Algoritmo BackPropagation - Gradiente descendente

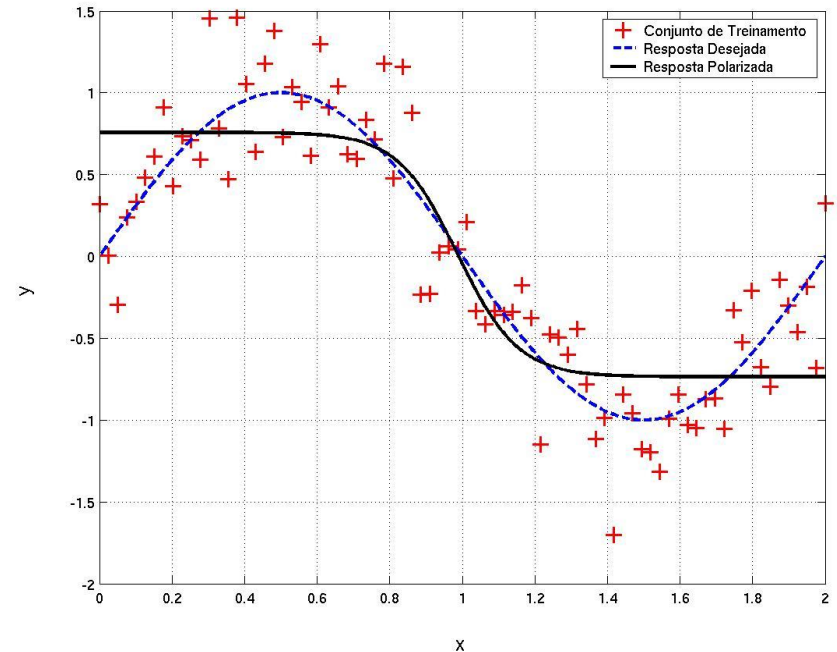
- É um método iterativo
- Os pesos são inicializados
- O ajuste é realizado segundo a direção do gradiente descendente da função de custo em relação ao vetor de pesos da rede

O Treinamento Supervisionado de RNAs

A Questão do Vício e da Variância



Overfitting
Variância



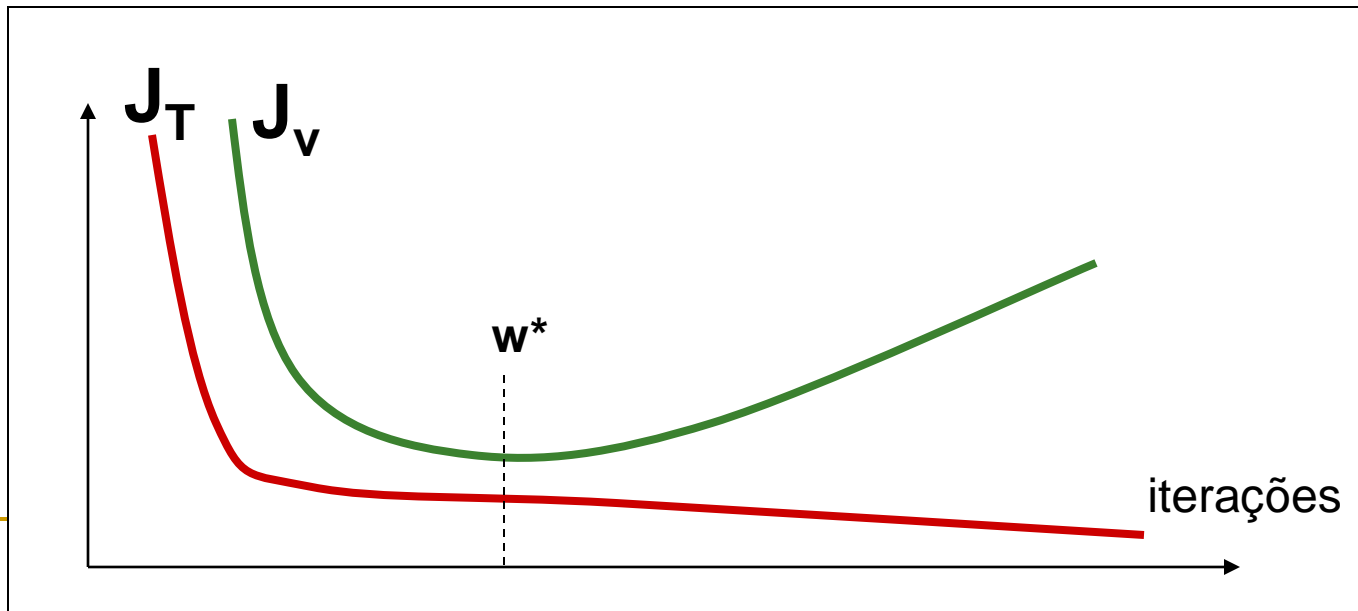
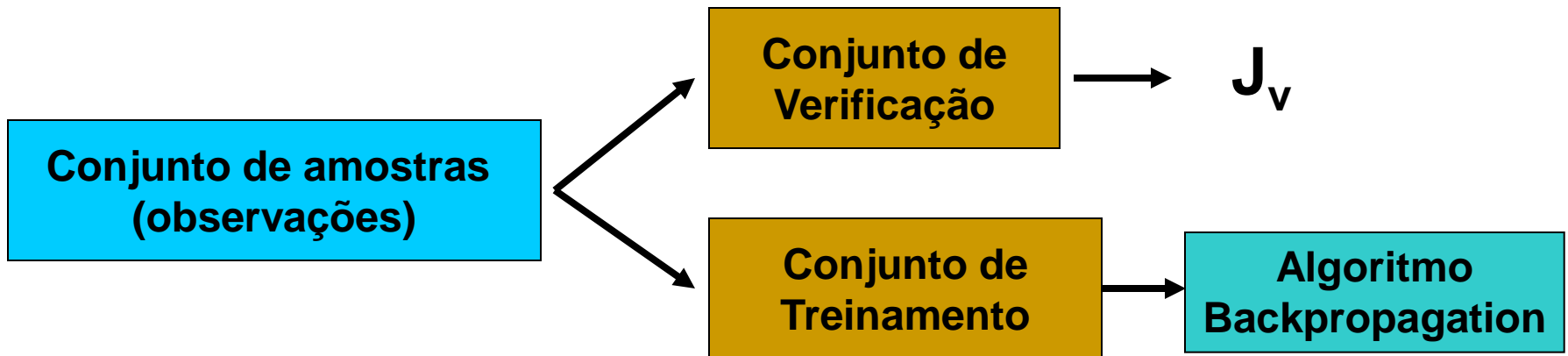
Underfitting
Vício

O Controle da Generalização

- Algoritmos de Validação [Amari et. al, 1996]
 - Cross-Validation [Stone, 1974]
 - Algoritmos de Pruning [Reed, 1993]
 - Regularização Bayesiana [Mackay, 1992]
 - Penalização (weight-decay [Hinton, 1989])
 - Multi-Objetivo [Teixeira, 2000, Costa 2002]
 - etc. ...
-

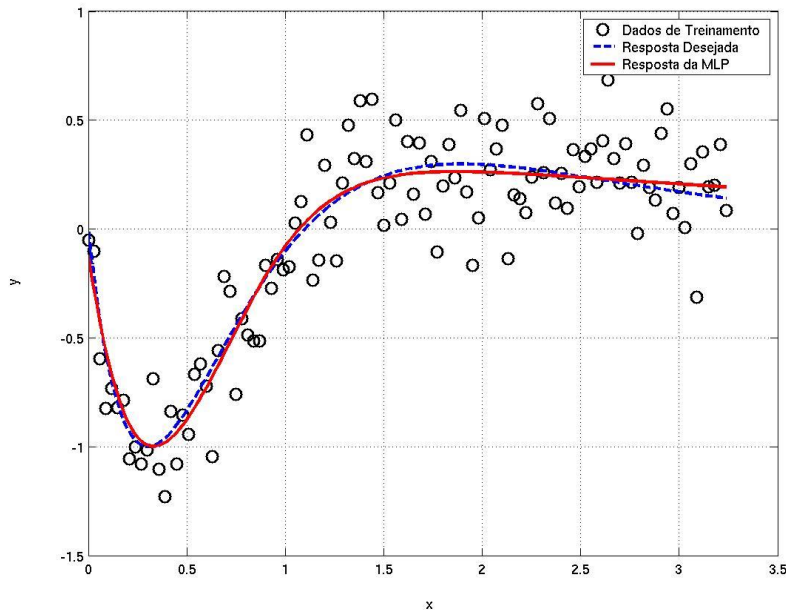
O Controle da Generalização

Early Stopping

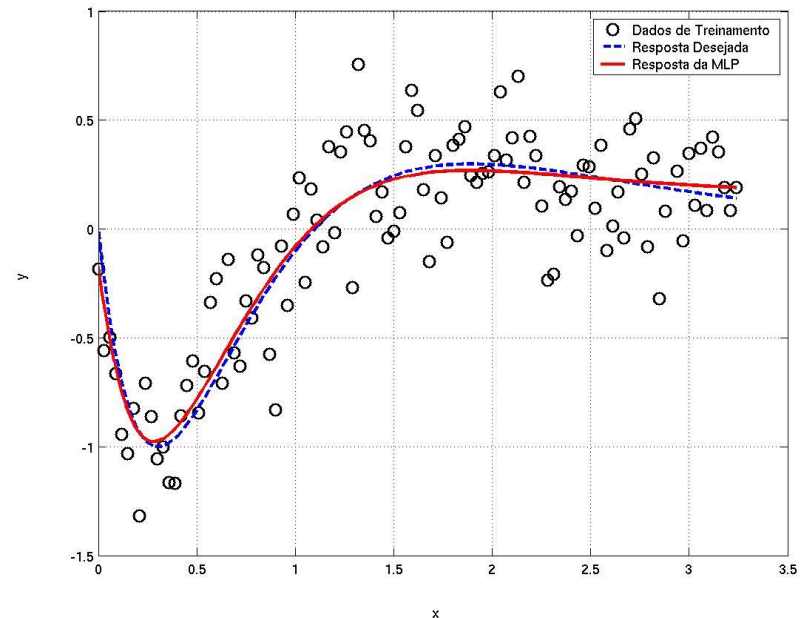


Exemplo de aproximação de função utilizando duas redes MLP com 12 nodos e 25 nodos na camada escondida, respectivamente.

O equilíbrio entre o erro e a norma garante modelos com alta capacidade de generalização independente do número de parâmetros (variáveis de ajuste) do modelo.

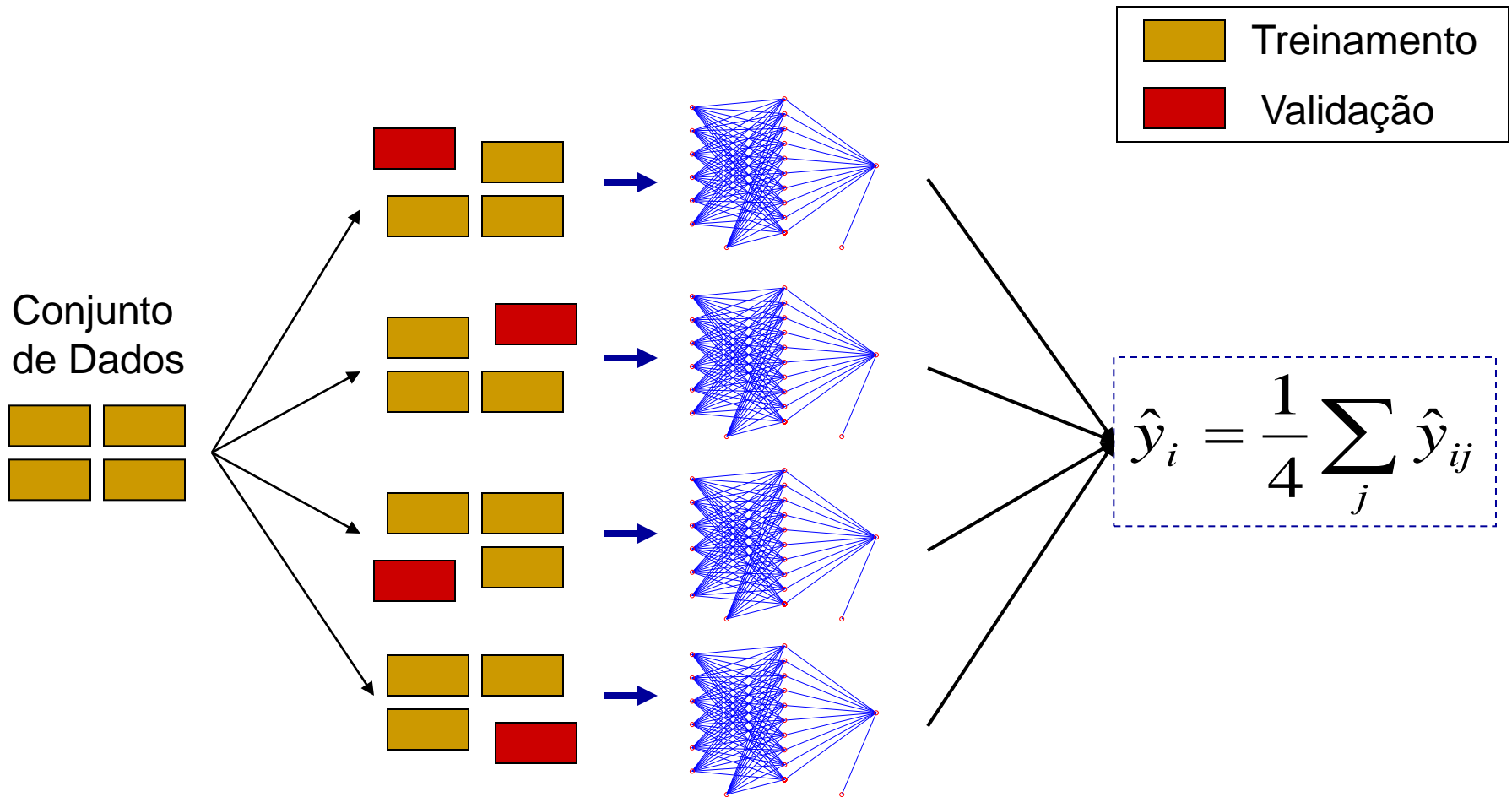


Rede MLP com 12 nodos na camada escondida



Rede MLP com 25 nodos na camada escondida

Validação Cruzada (*Cross-Validation*)

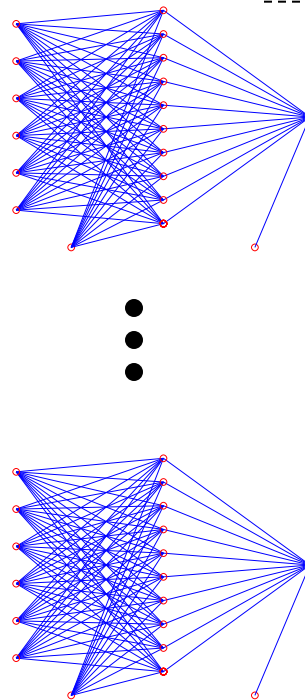
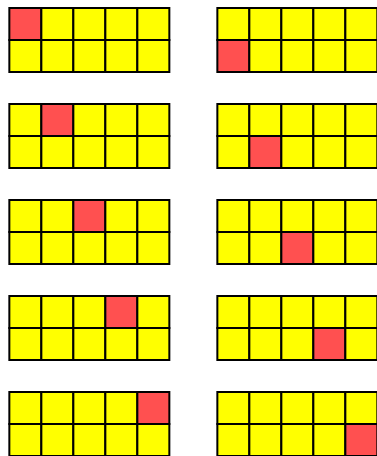


Multi-Layer Perceptron with 10-fold Cross-Validation

training	training	validation	training	training
training	training	training	training	training

The data is replicated 10 times, each replicate has a different validation subset. One MLP is trained for each partition using MOBJ-LASSO cross-validation

Data Replicates



$$\hat{y}_i = \frac{1}{10} \sum_j \hat{y}_{ij}$$

Aplicação das Redes Neurais Artificiais

- Aproximação de funções
 - Classificação de Padrões
 - Extração de conhecimento (data mining)
 - Previsão de Séries Temporais
 - *Clustering*
-

Topologia da RNA

